Supplemental Appendix for

Bespoke Realized Volatility: Tailored Measures of Risk for Volatility Prediction

by Andrew J. Patton and Haozhe Zhang

This version: May 11, 2025

This appendix presents additional details and analyses relevant to the main paper.

S.1. Simulation study

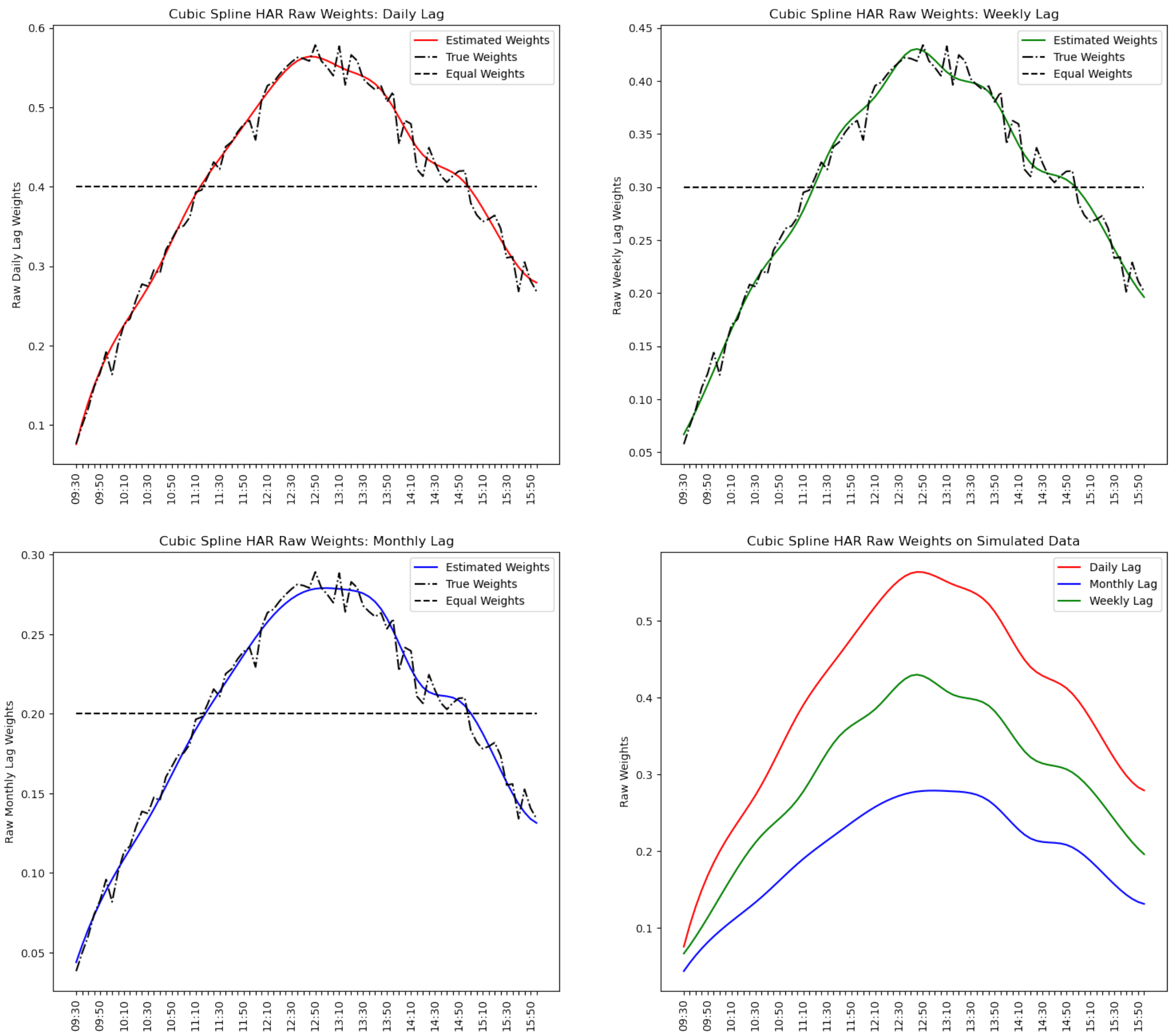
In this section we simulate data to illustrate the ability of the proposed Cubic HAR model to recover the true coefficients. We assume that high-frequency returns, $r_{i,t}$, are i.i.d. N(0,1/M), where M is the number of high-frequency returns each trade day. We combine these high frequency returns to obtain realized variances using time-of-day (TOD) weights, ω_i, as in equation 7 of the main paper, and we use a HAR model to simulate the time series dynamics of RV.

$$RV_t = \beta_0 + \beta_d \sum_{i=1}^{M} \omega_i r_{i,t-1}^2 + \beta_w \frac{1}{4} \sum_{j=2}^{5} \sum_{i=1}^{M} \omega_i r_{i,t-j}^2 + \beta_m \frac{1}{16} \sum_{j=6}^{21} \sum_{i=1}^{M} \omega_i r_{i,t-j}^2 + \epsilon_t \tag{S.1}$$

where $\epsilon_t \sim N(0,1)$. We set $\beta_0 = 0$, $\beta_d = 0.4$, $\beta_w = 0.3$, $\beta_m = 0.3$, and $M = 78$. We impose that ω_i follows the empirical TOD weights presented in Figure 4. Then we use the cubic HAR model with the exact same set-up (hyperparameter grid and estimation algorithm) to estimate the coefficients on the simulated data. Figure S.1 shows that when the true data generating process indeed is a linear combination with the TOD weights, our cubic HAR model will recover the true weights almost perfectly. (As the true weights in this simulation are not a smooth function of time, the cubic spline model clearly cannot obtain a perfect fit.) This simulation results give us confidence that the proposed Cubic

HAR has enough estimation power to recover the optimal weighting scheme for realized variance forecasting.

Figure S.1: Cubic HAR vs True Weights: Simulated Data



Note: This figure compares the estimated optimal weights (solid line) based on simulated data based with the true weights (dash-dotted line).

S.2. Data cleaning details

We use trade data from the NYSE Trade and Quote (TAQ) database. We follow Barndorff-Nielsen et al. (2009) to clean this data, using the following rules:

1. Keep only entries with time step between 9:30 am to 4:00 pm (when the exchange is open).
2. Delete entries with zero transaction prices.
3. Retain entries originating from NYSE and NASDAQ only.
4. Delete entries with corrected trades (CORR $\neq 0$).
5. Delete entries with unusual sale condition (COND with letter code except for letters E and F).
6. Use the median price if multiple entries have the same time stamp.

S.3. Alternative target for shrinkage

In the analysis in Section 3.2 of the main paper, we consider shrinkage methods (Ridge, LASSO and elastic net) that shrink the estimated parameters towards zero. This is a standard shrinkage target in high-dimensional estimation, but in our application an interesting alternative target is to shrink the parameters towards the benchmark HAR parameters. To do this, we re-write the regularized regression models as:

$$RV_t = \beta_0 + \widetilde{RV}_{t-1}(\beta^d_{HAR} + \boldsymbol{\gamma}^d) + \frac{1}{4}\sum_{j=2}^{5}\widetilde{RV}_{t-j}(\beta^w_{HAR} + \boldsymbol{\gamma}^w) + \frac{1}{16}\sum_{j=6}^{21}\widetilde{RV}_{t-j}(\beta^m_{HAR} + \boldsymbol{\gamma}^m) + e_t \tag{S.1}$$

That is, we split the coefficients on high frequency returns into the HAR coefficients and a perturbation term to be estimated. Then, instead of regularizing the entire coefficient, we only penalize the perturbation terms, and shrink these towards zero. That is, the penalty term becomes:

$$\alpha(\lambda\|\boldsymbol{\gamma}\|_1 + (1-\lambda)\|\boldsymbol{\gamma}\|_2) \tag{S.2}$$

Table S.1: Alternative regularized regression models: Shrink towards HAR

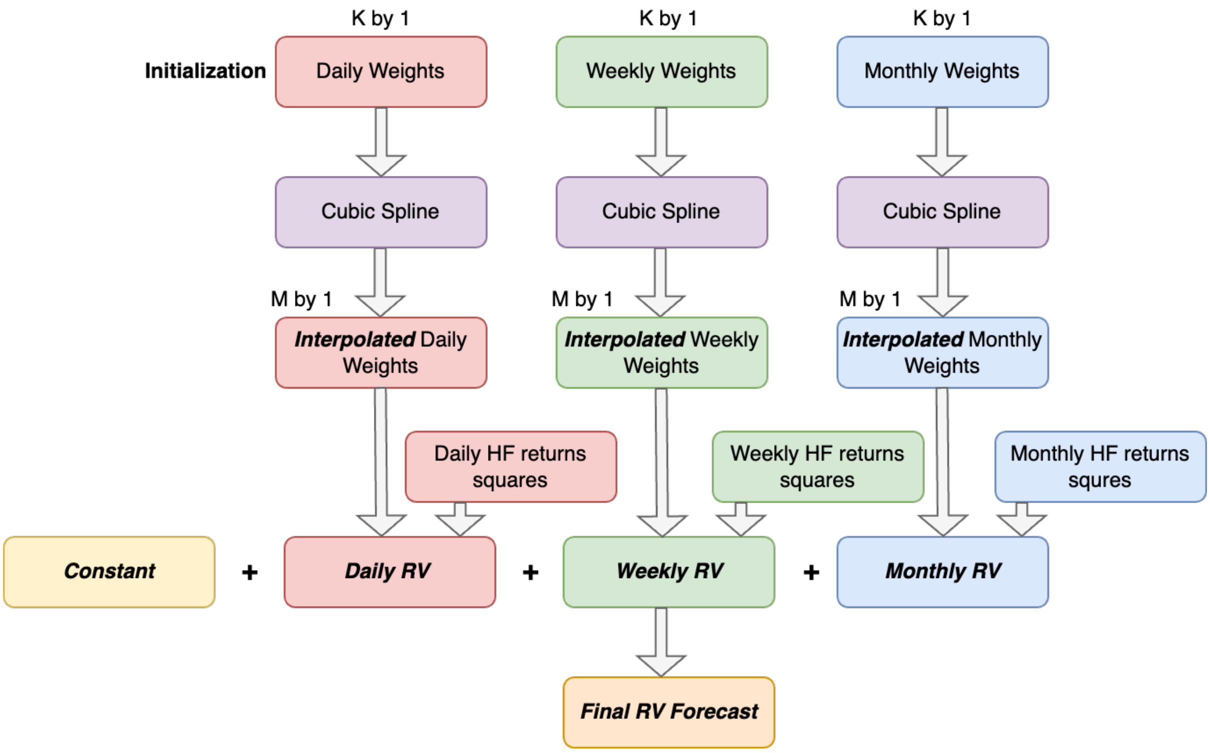
CubicHAR vs:	GW Losses Total	GW Losses Signif	GW Wins Total	GW Wins Signif	GW t-stat Panel
Ridge-Alt	166	44	720	423	-17.3
LASSO-Alt	165	46	721	437	-4.2
Elastic net-Alt	170	46	716	412	-16.4
OLS	54	17	832	564	-6.4
HAR	155	41	731	470	-21.7

Note: This table reports individual and panel Giacomini-White (2006) tests comparing the cubic HAR model against competing models across 886 S&P 500 stocks. A positive panel GW t-statistic indicates that the competing model out-performs the cubic HAR model, while a negative t-statstic indicates the opposite. The models labeled "-Alt" shrink the estimated coefficients towards the HAR model coefficients, rather than towards zero as in Table 2 of the main paper.

where $\boldsymbol{\gamma} = [\boldsymbol{\gamma}^d, \boldsymbol{\gamma}^w, \boldsymbol{\gamma}^m]$. The estimation procedure and hyperparameters grid are identical to those in the original regularized regression analysis. We compare the cubic HAR model with these alternative shrinkage estimators in Table S.1. (The OLS and HAR results are identical to those in Table 2 of the main paper, and are included here for ease of comparison.) With this alternative target the optimal degree of shrinkage is found to be large, and the Ridge, LASSO and elastic net models are all shrunk almost all the way to the benchmark HAR parameters. Given that, it is unsurprising that the cubic HAR continues to significantly out-perform these alternative regularized models, as shown in Table S.1, which have performance comparable to the benchmark HAR model.

S.4. Cubic HAR Model Architecture

Figure S.2: Cubic HAR Model Architecture

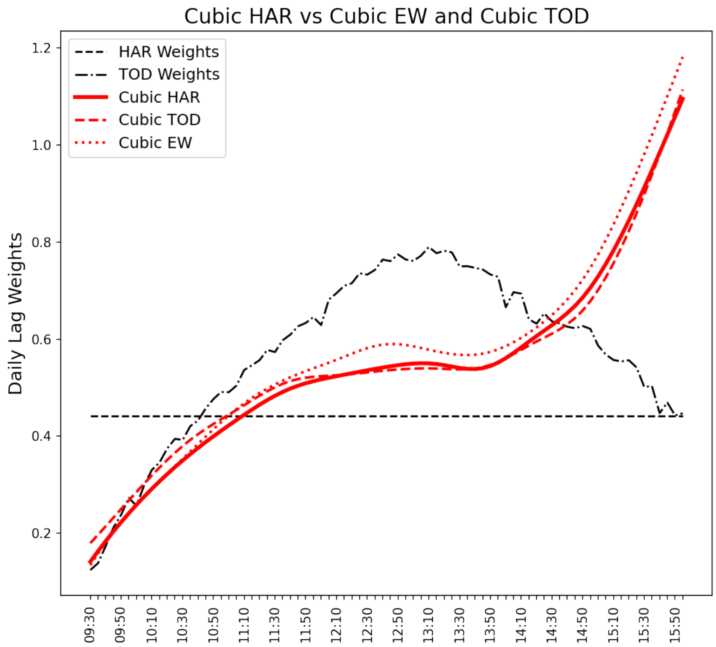


Note: This figure illustrates the cubic HAR model architecture. In particular, it shows how we use miniBatch gradient descent to iteratively solve the optimal parameters $(3 \times K + 1)$ and then use them to get the cubic spline interpolated weights for constructing final RV forecast. Here K represents the number of basis points for cubic spline interpolation or the true number of parameters, and M represents the desired number of points as cubic spline interpolation output.

S.5. Visualizing the optimal Cubic TOD and Cubic EW weights

This section presents the optimal daily weights for the cubic HAR, cubic TOD and cubic EW models, which differ in what shapes are estimated or imposed on the weekly and monthly weights. We observe the cubic TOD and cubic EW daily weights are almost identical to the cubic HAR weights for daily lags, which shows that restricting the shapes of weekly and monthly weights has almost zero affect on the estimated daily weights.

Figure S.3: Cubic TOD and Cubic EW weights



Note: This figure compares estimated cubic TOD and cubic EW weights against the cubic HAR model weights, averaged across the 886 S&P 500 stocks in our sample.

S.6. Additional tables

Table S.2: Comparisons involving log RV

Log Cubic vs:	GW Losses Total	GW Losses Signif	GW Wins Total	GW Wins Signif	GW t-stat Panel
Log HAR	148	59	738	563	-7.2
Cubic	812	733	74	29	19.3

Note: This table reports individual and panel Giacomini-White (2006) tests comparing the Cubic HAR model estimated on log RV against competing models across 886 S&P 500 stocks. The first row compares the model with HAR estimated on log RV. Both models are estimated using OLS and forecast comparisons are done using MSE. The second row compares forecasts of RV (non-logged) from the Cubic HAR model presented in Section 2.1 and the Cubic HAR model estimated on log RV, and then converted to a forecast of RV, using the adjustment in Clements and Preve (2021). These forecasts are evaluated using QLIKE loss. A positive panel GW t-statistic indicates that the competing model out-performs the log Cubic HAR model, while a negative t-statstic indicates the opposite.

Table S.3: HAR vs other models performance on S&P500. MSE loss.

HAR vs:	GW Losses Total	GW Losses Signif	GW Wins Total	GW Wins Signif	GW t-stat Panel
Fully Flexible	421	169	465	106	0.8
Flexible HAR	726	408	160	11	3.3
Cubic HAR	714	406	172	7	1.9
TOD HAR	653	348	233	62	2.2
Ridge	462	186	424	120	-0.1
LASSO	203	46	683	338	-4.1
Elastic net	434	166	452	136	-2.0
OLS	64	7	822	507	-3.3

Note: This table reports individual and panel Giacomini-White (2006) tests comparing the baseline HAR model against competing models across 886 S&P 500 stocks, using MSE loss. A positive panel GW t-statistic indicates that the competing model out-performs the HAR model, while a negative t-statstic indicates the opposite. This table is related to Table 1 of the main paper, which uses QLIKE loss.

Table S.4: Cubic HAR vs other models performance on S&P 500. MSE loss.

Cubic HAR vs:	GW Losses Total	GW Losses Signif	GW Wins Total	GW Wins Signif	GW t-stat Panel
Fully Flexible	151	16	735	383	-0.7
Flexible HAR	411	55	475	90	0.0
TOD HAR	405	69	481	196	-1.0
HAR	172	7	714	406	-1.9
Ridge	171	12	715	349	-1.7
LASSO	66	2	820	542	-5.2
Elastic net	159	8	727	372	-3.4
OLS	23	3	863	664	-3.3

Note: This table reports individual and panel Giacomini-White (2006) tests comparing the cubic HAR model against competing models across 886 S&P 500 stocks, using MSE loss. A positive panel GW t-statistic indicates that the competing model out-performs the cubic HAR model, while a negative t-statstic indicates the opposite. This table is related to Table 2 of the main paper, which uses QLIKE loss.

Table S.5: Multidays Ahead Volatility Forecasting: TOD HAR vs HAR

Horizon (days)	GW Losses Total	GW Losses Signif	GW Wins Total	GW Wins Signif	GW t-stat Panel
1	206	63	680	458	-18.6
2	253	72	633	349	-12.5
3	257	74	629	314	-10.1
4	264	74	622	293	-13.6
5	270	71	616	298	-11.5
20	253	68	633	345	-6.8
60	345	126	541	255	-0.2

Note: This table reports the individual and panel Giacomini-White (2006) tests results of TOD HAR model against the HAR model in the S&P500 cross section for longer horizon forecasting. Negative test statistics indicate the outperformance of the TOD HAR model.

Table S.6: Direct vs iterated multi-step forecasts

Horizon (days)	GW Losses Total	GW Losses Signif	GW Wins Total	GW Wins Signif	GW t-stat Panel
Panel A: Direct Cubic HAR vs Iterated HAR					
2	207	34	679	356	-12.9
3	259	41	627	290	-14.3
4	280	36	606	284	-9.2
5	258	31	628	274	-18.6
20	243	37	643	387	-25.2
60	159	51	727	576	-71.7
Panel B: Direct HAR vs Iterated HAR					
2	350	59	536	184	-6.4
3	410	78	476	152	-1.3
4	383	60	503	165	-5.5
5	356	45	530	186	-7.6
20	198	30	688	405	-29.2
60	133	43	753	626	-100.6

Note: This table compares the forecast accuracy of direct and iterated multi-step-ahead forecasts. "Direct" forecasts project h-step-ahead RV onto information available at time t. With the AR(21) structure embedded in the HAR model, it is also possible to generate "iterated" forecasts, which come from the model estimated for the one-step-ahead forecast and then iterated out to the horizon of interest, see Ghysels et al. (2019) for applications to volatility. Panel A compares direct forecasts from the Cubic HAR model with iterated forecasts from the HAR model, and is directly related to Panel A of Table 3 in the main paper. Similar to that table, we see here that direct forecasts from Cubic HAR significantly beat iterated forecasts from the HAR model. Panel B compares direct and iterated forecasts from the HAR model, and shows that iterated forecasts are worse for all horizons considered.

S.7. GARCH-X with RV as the target variable

The GARCH-X model in Section 4.2 of the main paper uses, effectively, the daily squared return as the volatility proxy when evaluating forecast accuracy. Realized variance is known to be a more accurate volatility proxy (see, e.g., Andersen and Bollerslev (1998) and Andersen et al. (2003)), and more accurate proxies lead to more powerful forecast comparisons, (see, e.g., Patton (2011)). Here we consider the estimating and evaluating the GARCH-X model replacing the squared daily return with 5-minute realized variance. Table S.7 presents results corresponding to Table 8 in the main paper. We see that GARCH-X based on bespoke RV continues to significantly outperform both the standard GARCH-X model, and the model use time-of-day (TOD) RV, with panel GW t-statistics less than -4 for both comparisons. The main difference between Table S.7 and Table 8 is that the number of individual GW tests that reject the null (listed in the "Signif" columns of the table) is greater, ranging from 68 to 186 in Table 8 and from 75 to 293 here. This increase in significance is consistent with 5-minute RV being a more accurate volatility proxy.

Table S.7: Bespoke RV for GARCH-X models, with RV as the target variable

	GW Losses		**GW Wins**		**GW t-stat**
Model	**Total**	**Signif**	**Total**	**Signif**	**Panel**
GARCH-X: Basic vs Bespoke	281	75	605	293	-2.6
GARCH-X: TOD vs Bespoke	528	151	358	84	-7.4

In Figure S.4 we plot the cross sectional average optimal weights for the bespoke GARCH-X model, estimated with 5-minute RV as the volatility proxy. This figure is similar to Figure 8 in the main paper.

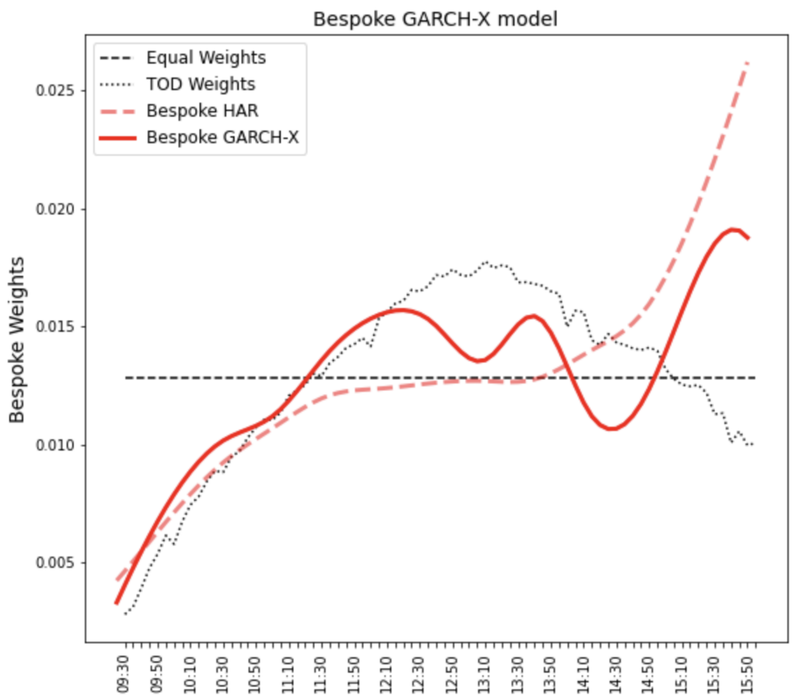
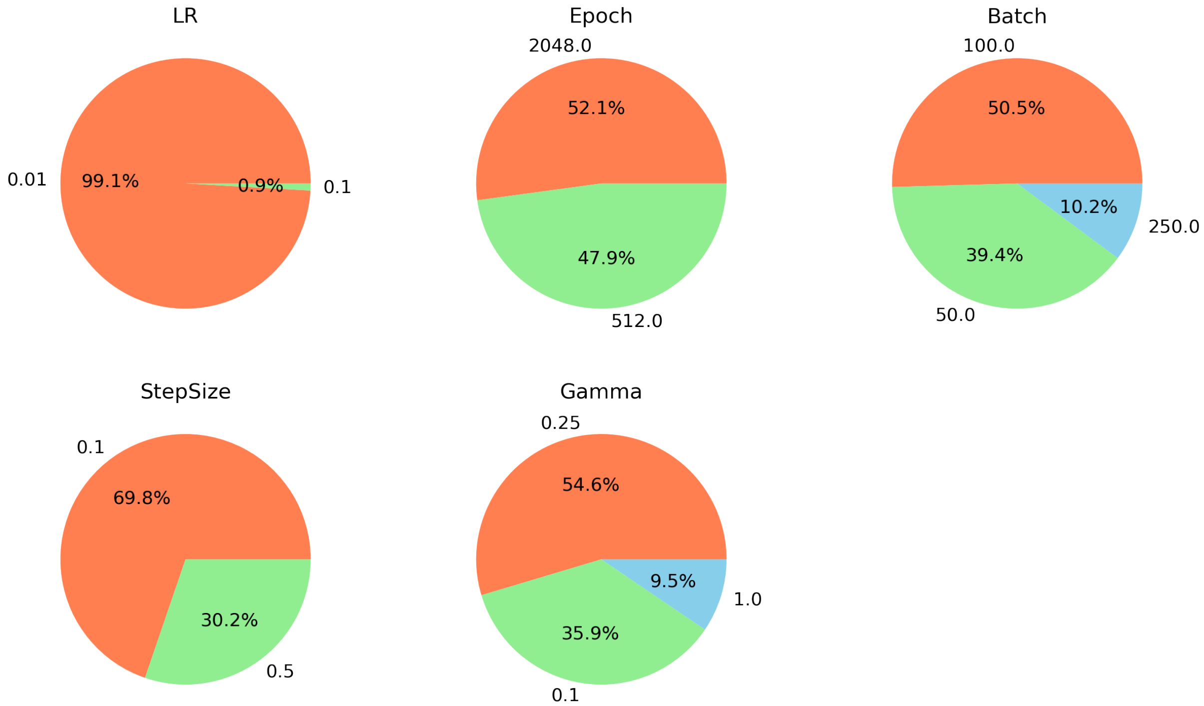


Figure S.4: Bespoke GARCH-X with RV target variable

S.8. Visualization of the hyper-parameters for deep learning based models

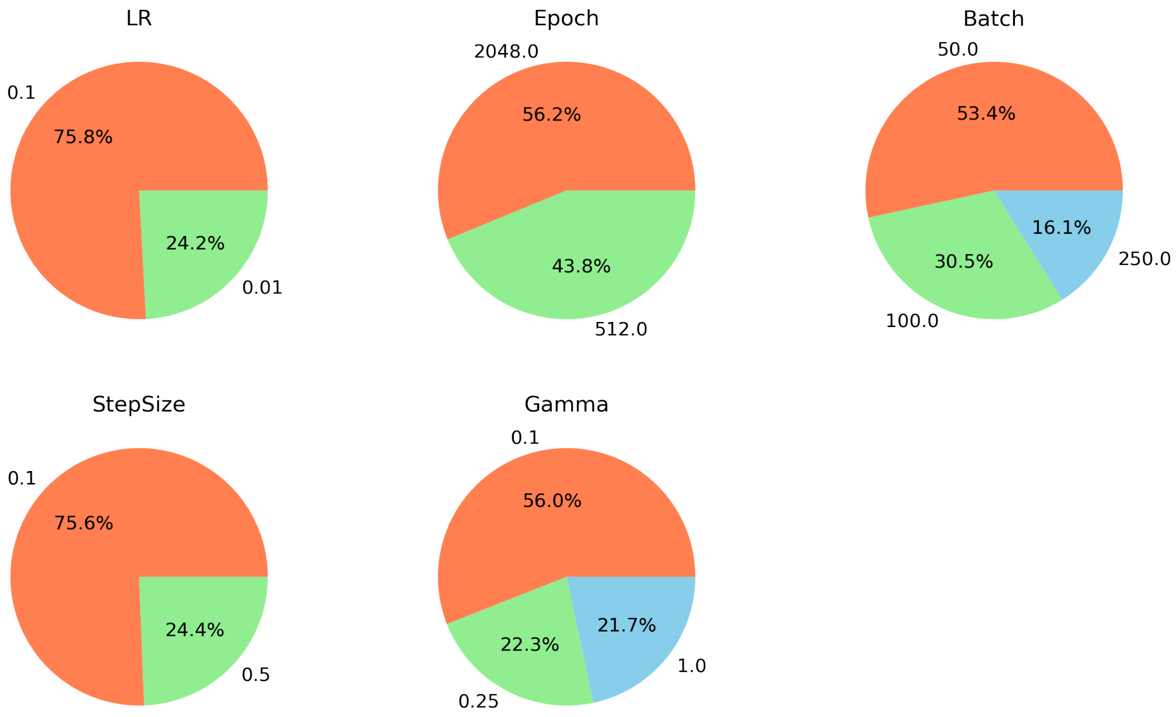
In this section, we also report visualization of the cross sectional variation in the deep learning based models.

Figure S.5: Fully flexible model hyperparameters



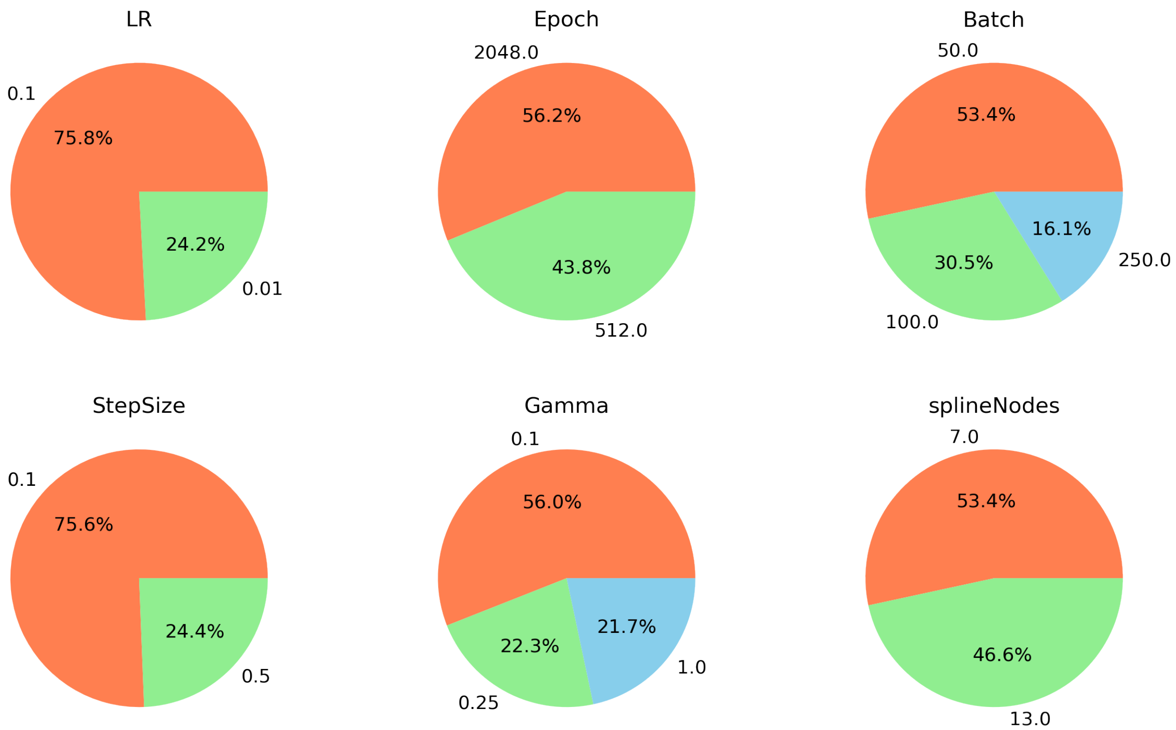
Note: This figure displays the cross-sectional variation of the optimal hyperparameters for the fully flexible model.

Figure S.6: Flexible HAR model hyperparameters



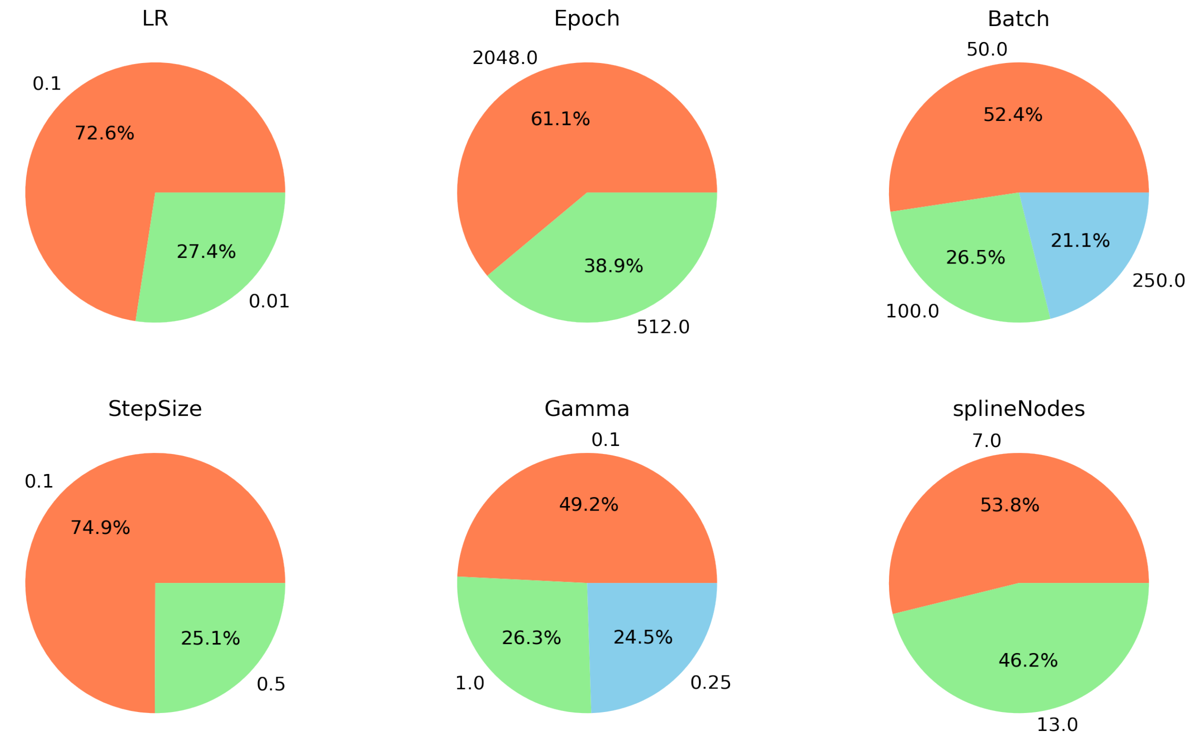
Note: This figure displays the cross-sectional variation of the optimal hyperparameters for the flexible HAR model.

Figure S.7: Cubic HAR model hyperparameters



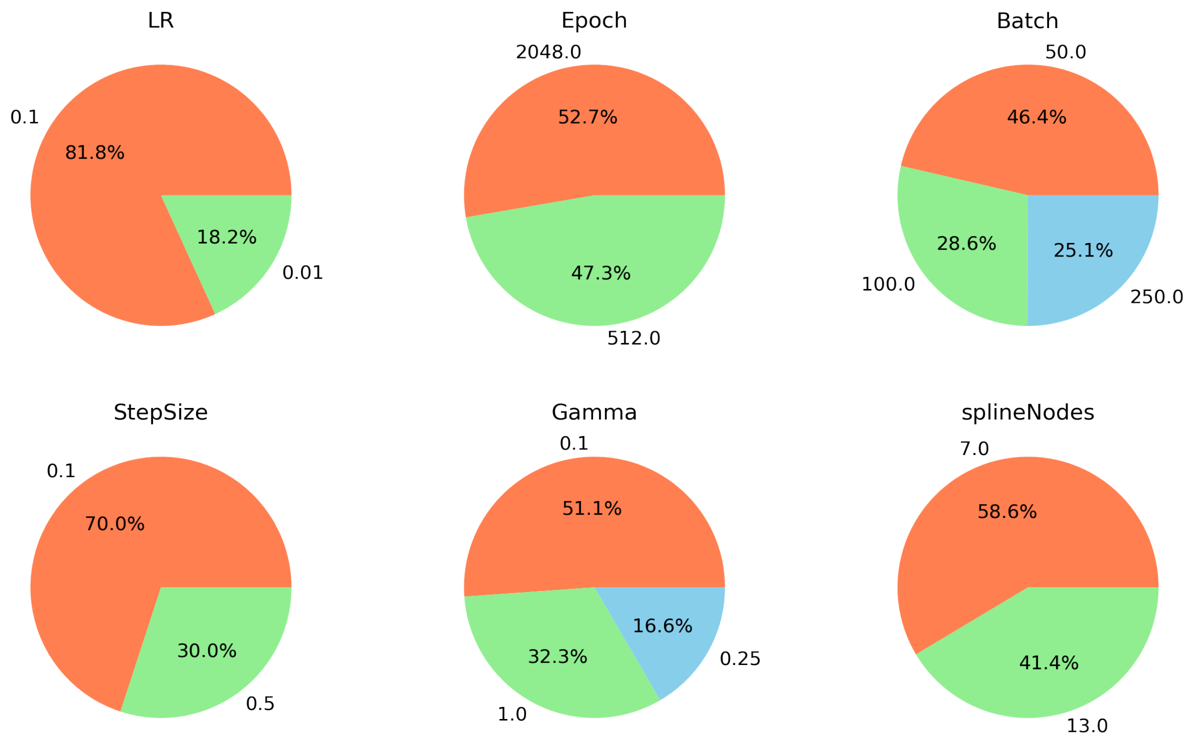
Note: This figure displays the cross-sectional variation of the optimal hyperparameters for the cubic HAR model.

Figure S.8: Cubic-TOD model hyperparameters



Note: This figure displays the cross-sectional variation of the optimal hyperparameters for the cubic-TOD model.

Figure S.9: Cubic-EW model hyperparameters



Note: This figure displays the cross-sectional variation of the optimal hyperparameters for the cubic-EW model.

Additional References

Andersen, T. G. and Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*, pages 885–905.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625.

Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., and Shephard, N. (2009). Realized kernels in practice: Trades and quotes. *Econometrics Journal*, 12(3):1–32.

Clements, A. and Preve, D. P. A. (2021). A practical guide to harnessing the HAR volatility model. *Journal of Banking & Finance*, 133:106285.

Ghysels, E., Plazzi, A., Valkanov, R., Rubia, A., and Dossani, A. (2019). Direct versus iterated multiperiod volatility forecasts. *Annual Review of Financial Economics*, 11(1):173–195.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1):246–256.